

## Converting text input into moving-face picture

Patent Number:  GB2231246

Publication date: 1990-11-07

Inventor(s): HIGUCHI NORIO;; KOIKE ATSUSHI;; KANEKO MASAHIKE;; HATORI YOSHINORI;; YAMAMOTO SEIICHI

Applicant(s): KOKUSAI DENSHIN DENWA CO LTD (JP)

Requested Patent:  JP2234285

Application Number: GB19900005142 19900307

Priority Number(s): JP19890053899 19890308

IPC Classification: G06F15/72

EC Classification: G10L21/06L, G06T9/00F, G06T15/70, G10L9/20, H04N7/26J4

Equivalents: JP2518683B2

### Abstract

A moving picture of a face with mouth-shape variations corresponding to a text sentence input is produced. The input sentence is divided into a train of phonemes and a speech synthesis technique capable of outputting a voice feature of each phoneme and its duration is utilized. Based on the voice feature, a mouth-shape feature corresponding to each phoneme is determined 3. Based on the mouth-shape feature, the value of a mouth-shape parameter is determined 5, 4 for representing a mouth shape. Further, the value of the mouth-shape parameter for each frame of the moving picture is controlled 2 in accordance with the duration of each phoneme, thereby synthesizing the moving face picture having mouth-shape variations which agree with the speech output. 

Data supplied from the **esp@cenet** database - 12

## ⑫ 公開特許公報 (A)

平2-234285

⑬ Int. Cl.

G 06 F 15/72  
G 10 L 3/00  
9/20

識別記号

350  
S  
A

庁内整理番号

8125-5B  
8622-5D  
8622-5D

⑭ 公開 平成2年(1990)9月17日

審査請求 未請求 請求項の数 3 (全9頁)

## ⑮ 発明の名称 画像合成方法及びその装置

⑯ 特願 平1-53899

⑯ 出願 平1(1989)3月8日

⑰ 発明者 金子 正秀 東京都新宿区西新宿2丁目3番2号 国際電信電話株式会社内

⑰ 発明者 小池 淳 東京都新宿区西新宿2丁目3番2号 国際電信電話株式会社内

⑰ 発明者 羽鳥 好律 東京都新宿区西新宿2丁目3番2号 国際電信電話株式会社内

⑰ 発明者 山本 誠一 東京都新宿区西新宿2丁目3番2号 国際電信電話株式会社内

⑯ 出願人 国際電信電話株式会社

⑯ 代理人 弁理士 大塚 学 外1名

最終頁に続く

## 明細書

## 1. 発明の名称

画像合成方法及びその装置

## 2. 特許請求の範囲

(1) 文字列として表現される文章を入力し、これに対応した口形状変化を有する顎動画像を生成する画像合成方法において、前記文字列を音素列に分割し、各音素ごとに音声特徴及び持続時間を出力することが可能な音声合成手法を利用し、音声特徴に基づいて各音素に対応する口形特徴を決定し、更に該口形特徴に従って具体的な口形状を表現するための口形状パラメータの値を決定し、また、各音素ごとの該口形状パラメータの値に対して前記各音素ごとの持続時間に基づいて動画像の各フレームごとに与えられる口形状パラメータの値を制御し、音声出力に適合した口形状変化を有する顎動画像の合成を行うことを特徴とする画像合成方法。

(2) 文字列として表現される文章を入力するための入力端子と、該入力端子より入力される該文字列を音素列に分割し、各音素ごとに音声特徴及び持続時間を出力することが可能な音声合成部と、各音素ごとの該音声特徴から口形特徴への変換を行う変換部と、種々の口形特徴と具体的な口形状を表現する口形状パラメータとを対応付けた変換テーブルと、前記変換部で得られる各音素ごとの口形特徴に対応する口形状パラメータを前記変換テーブルから取出す口形状パラメータ取得部と、一定時間間隔の画像系列として与えられる動画像を生成するために該口形状パラメータ取得部から得られる口形状パラメータの値の出力を前記音声合成部から与えられる各音素ごとの持続時間に従って制御するための時間調整部と、該時間調整部の制御のもとに前記口形状パラメータ取得部から出力される口形状パラメータの値に従って画像を生成するための画像生成部とを備えたことを特徴とする画像合成装置。

(3) 前記時間調整部の出力に従って、ある音素から次の音素への遷移を検出するための遷移検出部と、前記画像生成部で用いられる口形状パラメータの値を少なくとも1フレーム時間以上保持することが可能なメモリと、該メモリに保持されている口形状パラメータの値と前記口形状パラメータ取得部より与えられる口形状パラメータの値との中間値を求める口形状パラメータ修正部とを更に備え、ある音素から次の音素への遷移時に中間的な口形状を生成して滑らかな口形状変化を有する顔動画像を生成することを特徴とする特許請求の範囲第2項記載の画像合成装置。

### 3. 発明の詳細な説明

#### (発明の技術分野)

本発明は、ディジタル処理による画像合成方法に関するもので、特に、発声に伴う口形状変化を表現する顔画像（静止画像または動画像）を合成する方式に関するものである。

#### (従来技術)

人が発声する場合には、調音器官により音声情報が生成され、同時に、外見的な変化として発声に伴い、口部分の動き（形状変化）が生じる。人が直接発声するのではなく、文字列として入力された文章を音声情報に変換して出力する方法は音声合成と言われ、従来、多くの成果が得られてきている。一方、入力された文章に対して対応する口形状変化を有する顔画像を生成する方法に関しては従来技術は少なく、松岡清利、黒須顕二による次の報告があるにどどまっている。

松岡、黒須の方法は、〔松岡清利、黒須顕二：「聴覚障害者の発話訓練のための動画プログラム」電子情報通信学会論文誌、vol.J70-D,no.11,PP.2167-2171（1987年11月）〕に示されている。これは、プログラムの形で実施されているが、入力された文章に対して、対応する口形状変化を得るために考え方の基本を整理して示すと、第6図のようになる。

第6図において、50は音節分離部、51は音節と

口形パターンの対応付け部、52は音節と口形パターンの対応テーブル、53は口形状選択部、54は口形状用メモリである。次に各部の動作を簡単に説明する。音節分離部50は、入力された文章（文字列）に対して、これを音節単位に区切る働きをする。例えば「kuma」という入力は、「ku」と「ma」の2つの音節に分けられる。次に、音節と口形パターンの対応テーブル52は、予め用意された音節と口形パターンの対応関係を蓄積したテーブルである。音節は“a”、“ka”などひとまとまりの音を表現するものである。口形パターンは、大口形(< A > < I > < U > < E > < K > 等)と小口形(< u > < o > < k > < s > 等)とがあり、口形の種類を示すものである。これらを用いて“a”に対しては< A > < \* > < A >，“ka”に対しては< K > < \* > < A > というように音節と口形パターンの対応関係をテーブルにしておくわけである。ここで、< \* >は中間口形を示す。音節と口形パターンの対応付け部51では、音節分離部50から送られてくる一つ一つの音節ごとに、音節と口

形パターンの対応テーブル52を参照して、対応する口形パターンをテーブルから読出す。次に口形状用メモリ54は、前述の口形パターンの各々について具体的な口形状を図形或いは形状パラメータの形で蓄積したメモリである。口形状選択部53では、音節と口形パターンの対応付け部51から送られてくる口形パターン列に対して、順次口形状用メモリ54を参照して、具体的な口形状を選択し、画像として出力する。この時、必要に応じて中間形状（前後の口形状の中間の形状）の生成も行われる。なお、動画像としての出力のために、各音節に対して固定的に4フレーム分の口形状を生成するようになっている。

この外に、関連する従来技術として、文章入力に対してではないが、音声を入力として対応する口形状変化を推定する方法も報告されている。これは、〔森島繁生、相沢清晴、原島博：「音声情報に基づく表情の自動合成の研究」第4回NICOGRAPH論文コンテスト論文集、PP.139-146、日本コンピュータ・グラフィックス協会（1988年11月）〕に

示されている。ここでは、入力された音声情報に対して、対数平均パワーを計算して口の開き具合を制御する方法と、声道のホルマント特徴に対応する線形予測係数を計算して口形状を推定する方法と2通りが提案されている。

(発明が解決しようとする課題)

従来技術として、文章(文字列)を入力して、これに対応する口形状変化を有する顔画像を生成するための方法については、松岡、黒須の方法を示したが、次のような点で問題がある。すなわち、発声においては音声出力と口形状とに密接な関係があるにもかかわらず、基本的には文章を文節に区切って文字上の対応から口形パターンを選択しており、音声生成の機構と口形状生成との関連付けが不十分である。従って、音声出力との的確に対応した口形状生成が困難であるという問題がある。次に、音素(発声における最小単位、音節は複数の音素の組合せからなる)については、前後の音素とのつながり等によって、持続時間が異なるにもかかわらず、松岡、黒須の方法では、各音節に

固定的に4フレームを割当てており、入力される文章に応じた自然な口形状変化を表現することが困難であるという問題がある。また、入力された文章に対して、音声と口形状画像を同一のタイミングで出力しようとした場合に両者でのマッチングをとることも困難である。

さらに、森島、相沢、原島の方法は、入力された音声情報をもとにして口形状を推定するという技術であり、文章を入力して、これに対応した口形状変化を有する動画像を生成するという目的には適用することができない。

(発明の目的)

本発明は、上述した従来技術の問題点を解決するためになされたのもであり、音声出力との的確な対応付けがなされ、かつ、各音素の持続時間に合わせた形で口形状変化を表現することが可能な画像合成方法及びその装置を提供することを目的とする。

(発明の構成)

本発明の第1の特徴は、文字列として表現され

る文章を入力し、これに対応した口形状変化を有する顔動画像を生成する画像合成方法において、前記文字列を音素列に分割し、各音素ごとに音声特徴及び持続時間を出力することが可能な音声合成手法を利用し、音声特徴に基づいて各音素に対応する口形特徴を決定し、更に該口形特徴に従って具体的な口形状を表現するための口形状パラメータの値を決定し、また、各音素ごとの該口形状パラメータの値に対して前記各音素ごとの持続時間に基づいて動画像の各フレームごとに与えられる口形状パラメータの値を制御し、音声出力に適合した口形状変化を有する顔動画像の合成を行うことにある。

本発明の第2の特徴は、文字列として表現される文章を入力するための入力端子と、該入力端子より入力される該文字列を音素列に分割し、各音素ごとに音声特徴及び持続時間を出力することが可能な音声合成部と、各音素ごとの該音声特徴から口形特徴への変換を行う変換部と、種々の口形特徴と具体的な口形状を表現する口形状パラメー

タとを対応付けた変換テーブルと、前記変換部で得られる各音素ごとの口形特徴に対応する口形状パラメータを前記変換テーブルから取出す口形状パラメータ取得部と、一定時間間隔の画像系列として与えられる動画像を生成するために該口形状パラメータ取得部から得られる口形状パラメータの値の出力を前記音声合成部から与えられる各音素ごとの持続時間に従って制御するための時間調整部と、該時間調整部の制御のもとに前記口形状パラメータ取得部から出力される口形状パラメータの値に従って画像を生成するための画像生成部とを備えたことにある。

本発明の第3の特徴は、文字列として表現される文章を入力するための入力端子と、該入力端子より入力される該文字列を音素列に分割し、各音素ごとに音声特徴及び持続時間を出力することが可能な音声合成部と、各音素ごとの該音声特徴から口形特徴への変換を行う変換部と、種々の口形特徴と具体的な口形状を表現する口形状パラメータとを対応付けた変換テーブルと、前記変換部で

得られる各音素ごとの口形特徴に対応する口形状パラメータを前記変換テーブルから取出す口形状パラメータ取得部と、一定時間間隔の画像系列として与えられる動画像を生成するために該口形状パラメータ取得部から得られる口形状パラメータの値の出力を前記音声合成部から与えられる各音素ごとの持続時間に従って制御するための時間調整部と、該時間調整部の制御のもとに前記口形状パラメータ取得部から出力される口形状パラメータの値に従って画像を生成するための画像生成部とをに加えて、前記時間調整部の出力に従ってある音素から次の音素への遷移を検出するための遷移検出部と、前記画像生成部で用いられる口形状パラメータの値を少なくとも1フレーム時間以上保持することが可能なメモリと、該メモリに保持されている口形状パラメータの値と前記口形状パラメータ取得部より与えられる口形状パラメータの値との中間値を求める口形状パラメータ修正部とを更に備え、ある音素から次の音素への遷移時に中間的な口形状を生成して滑らかな口形状変化

を有する顔動画像を生成することにある。

(実施例1)

第1図は、本発明における第1の実施例を説明するためのブロック図である。入力情報としては、キーボード或いは磁気ディスク等のファイル装置から得られる文字列(文章)を考える。第1図において、1は音声合成部、2は時間調整部、3は音声特徴から口形特徴への変換部、4は口形特徴から口形状パラメータへの変換テーブル、5は口形状パラメータ取得部、6は画像生成部、10はゲート、900は文字列入力用の端子、901は画像出力用の端子である。

次に各部の動作について説明する。音声合成部1は入力された文字列に対応した音声出力を合成する部分である。音声合成に関しては従来各種の方式が提案されているが、ここでは、口形状生成との整合性が優れているという点から、声道モデルとしてKlatt型ホルマント音声合成器を用いた既存の音声規則合成手法の利用を想定している。この手法に関しては、(山本誠一、樋口直男、清

水徹：「テキスト編集機能付き音声規則合成装置の試作」電子情報通信学会技術報告SP87-137(1988年3月)に詳しく述べられている。音声合成部そのものは既存技術であり、また本発明が目的とする部分ではないので詳細な説明は省略する。但し、音声生成と口形状との的確な対応をとるために、各音素毎に音韻特徴及び持続時間に関する情報が出力されることが必要である。山本、樋口、清水の手法では、調音様式、調音点、有声/無声の区別、ピッチ制御情報などの音韻特徴及びこれに基づく持続時間の情報が出力されるようになっており、この要求を満足している。これらの情報が得られるのであれば、他の音声合成方法を利用するものであっても差し支えはない。

次に時間調整部2は、音声合成部1より得られる各音素毎の持続時間(第1番目の音素の持続時間をもとする)に基づいて、画像生成部6への口形状パラメータの受渡しを制御するためのものである。すなわち、テレビジョン信号として画像(特に動画像)を出力するためには、例えばNTSC方

式の場合毎秒30フレーム(1フレーム当たり1/30秒)であり、1/30秒毎の情報に直して画像を生成する必要がある。時間調整部2の詳しい動作については後述する。

次に、音韻特徴から口形特徴への変換部3では、音声合成部1から得られる音韻特徴に基づいて、該当音素に対応する口形特徴への変換を行う。口形特徴としては、例えば、(1)口の開き具合(かなり開いている～完全に閉じている)、(2)唇の丸め具合(丸めている～横に引いている)、(3)下顎の高さ(上がっている～下がっている)、(4)舌の見え具合、を考える。各種の音素に対して、人間が実際にどう発声しているかに関する観察に基づいて、音韻特徴と口形特徴との対応を規則化している。

例えば、“konnichiwa”という文章が入力された場合、

井井(無声)	lv0	lh4	jaw0
k	lv2	lh2	lback
o	lv2	lh1	jaw2
:			

のような形で口形特徴への変換がなされる。ここで、 $l_v, l_h, l_w$  は各々口の開き具合、唇の丸め具合、下顎の高さを示しており、数字は程度を表している。 $x$  は程度が前後の音素によって決められることを示している。また、 $l_{bck}$  は舌の見え具合を示している（この場合、舌が奥の方にわずかに見えることを表している）。

口形特徴から口形状パラメータへの変換テーブル 4 は、音声特徴から口形特徴への変換部 3 で得られる前述の口形特徴の各々について、具体的な口形状を表現するためのパラメータの値を与えるテーブルである。ここで第 2 図は、口形状を表現するためのパラメータの例を示した図である。第 2 図(a)は口部分を正面から眺めた時の正面図であり、点  $P_1 \sim P_8$  の 8 点の位置により口形状を、点  $Q_1, Q_2$  の位置により上、下の歯の見え具合を、 $h_1, h_2$  の値により上、下の唇の厚みを与える。第 2 図(b)は口部分を横から眺めた時の側面図であり、 $\theta_1, \theta_2$  の角度により、上、下の唇のめくれを与える。変換テーブル 4 では、前述の口形特

徴の各々について、実際に人が発声をする時の口形状に対する計測結果を参考にして前もって定められた上記パラメータ  $P_1 \sim P_8, Q_1 \sim Q_2, h_1, h_2, \theta_1, \theta_2$  の値の組をテーブルの形で保持しておく。

口形状パラメータ取得部 5 では、音声特徴から口形特徴への変換部 3 より得られる該当音素に対する口形特徴に対して、口形特徴から口形状パラメータへの変換テーブル 4 を参照して、該当音素に対する口形状パラメータの値の組を取得する。

ゲート 10 は、該当音素に対する上記口形状パラメータを画像生成部 6 に送るか否かを制御するためのものであり、時間調整部 2 から指示された回数（この回数に  $1/30$  秒を乗じた値が、該当音素に対する口形状の表示時間となる）だけ、上記口形状パラメータを画像生成部 6 に送る。

画像生成部 6 はゲート 10 を介して口形状パラメータ取得部 5 より送られてくる  $1/30$  秒毎の口形状パラメータに基づいて口形状画像の生成を行う。必要に応じて顔全体を含めた画像の生成を行う。

口形状パラメータを与えての口形状画像ないし顔画像の生成に関する詳細については、例えば〔金子正秀、羽鳥好律、小池淳：「形状変化の検出と 3 次元形状モデルに基づく顔動画像の符号化」電子情報通信学会論文誌 B、vol. J71-B, no. 12, pp. 1554-1563(1988年12月)〕に述べられている。概略としては、人物頭部の 3 次元形状を表現する 3 次元ワイヤフレームモデルを予め用意しておく。与えられた口形状パラメータに従って 3 次元ワイヤフレームモデルの口部分（具体的には、唇、歯、顎等）の形状を変形する。この変形後のモデルに、各部の濃淡や色を表現する情報を画素単位で付与することにより、リアルな口形状画像ないし顔画像を得ることができる。

ここで、時間調整部 2 の動作について詳しく説明する。第 3 図は時間調整部 2 の動作を説明するためのブロック図である。第 3 図において、21 は遅延部、22 は大小判定部、23, 24 はメモリ、25, 26 は加算器、27 はスイッチ、28, 29 は分岐、30 は時間正規化部、201, 202 は大小判定部 22 の出力

線、902 は初期リセット用の端子、903 は定数( $1/30$ )入力用端子、920, 921 はスイッチ 27 に関わる端子である。次に各部の動作について説明する。メモリ 23 は、1 番目の音素までの合計の持続時間  $\sum_{i=1}^{1-1} t_i$  を蓄えておくためのメモリである。画像合成を始める前に、端子 902 より与えられる初期リセット信号で蓄がセットされる。音声合成部 1 から 1 番目の音素の持続時間が与えらるると、加算器 25 により、メモリ 23 に蓄えられた  $1 - 1$  番目の音素までの合計の持続時間  $\sum_{i=1}^{1-1} t_i$  と  $t_i$  の和  $\sum_{i=1}^{1-1} t_i + t_i$  が求められる。遅延部 21 は、 $1 - 1$  番目の音素までの合計の持続時間  $\sum_{i=1}^{1-1} t_i$  を、 $1 + 1$  番目の音素に対する処理に入るまで蓄積する働きをする。時間正規化部 30 では、遅延部 21 の出力  $\sum_{i=1}^{1-1} t_i$  に対し、 $1/30 \times N \leq \sum_{i=1}^{1-1} t_i < 1/30 \times (N+1)$  を満足する  $N$  を求め、 $1/30 \times N$  の値を出力する。ここで、 $N$  は整数、また、 $1/30$  は 1 フレームの時間  $1/30$  秒を与える定数である。スイッチ 27 は、1 番目の音素に対する処理に入る時に、大小判定部 22 からの出力線 202 により端子 920 の側に接続される。この時、

加算器26により、時間正規化部30の出力 $1/30 \times N$ と定数 $1/30$ との和 $t$ が計算される。大小判定部22では、この $t$ の値と $\sum t_i$ の値との大小を比較し、 $t \leq \sum t_i$ の場合には、出力線201に、また $t > \sum t_i$ の場合には出力線202に信号を出力する。 $t > \sum t_i$ の場合は、1番目の音素の持続時間が終了したことを意味し、出力線202を介して、音声合成部1へ1+1番目の音素に関する情報を出力するための指示、メモリ24へ内容をリセットするための指示、スイッチ27へ端子920へ接続するための指示、遅延部21へ遅延されていた $t_i$ の値を出力するための指示がなされる。メモリ24は、加算器26の出力を一時的に蓄えておくためのものである。スイッチ27は $t \leq \sum t_i$ が成立つ間端子921に接続されており、加算器26により、順次、今までの $t_i$ に $1/30$ を加えたものを新た $t$ にする操作が行われる。以上により、 $t \leq \sum t_i$ が成立つ間、大小判定部22より出力線201に信号が出力され、この信号により第1図におけるゲート10が制御されることにより、1番目の音素の持続時間

説明する。

遷移検出部8は、ある音素（例えば1番目の音素）から次の音素（1+1番目の音素）への遷移を検出するためのものである。第5図は本発明による遷移検出部8の動作を説明するためのブロック図であり、81はカウンタ、82は判定回路、210、211は出力線である。カウンタ81は、大小判定部22からの出力線202に信号が出力された時に0にリセットされる。また、大小判定部22において出力線201に信号が出力されるごとに1ずつカウントアップする。判定回路82では、カウンタ81の出力が“1”であるか否かを判定し、“1”的には、ある音素から次の音素への遷移が生じたということであるので、出力線210に信号を出力する。一方、2以上の時には、現在の音素が持続しているということであるので、出力線211に信号を出力する。

メモリ9は、前フレームの画像を合成するために用いられた口形状パラメータを少なくとも1フレーム期間蓄えておくためのメモリである。口形

の間、1番目の音素に対する口形状パラメータが画像生成部6に供給される。

以上が本発明の第1の実施例に対する説明である。ここで、第1の実施例の場合、1番目の音素から1+1番目の音素に移る場合、1番目の音素に対する口形状パラメータから、1+1番目の音素に対する口形状パラメータへと不連続に変化することになる。両者の口形状パラメータに極端な違いがなければ、合成される動画像には余り不自然さは生じない。しかし、人間が発声をする場合、口形状は連続的に変化しており、1番目の音素から1+1番目の音素に移る場合、口形状が連続的に変化することが望ましい。

#### （実施例2）

第4図はこの要求を満足するための本発明の第2の実施例を説明するためのブロック図である。第4図において、7は口形状パラメータ修正部、8は遷移検出部、9はメモリ、40はスイッチ、910、911はスイッチ40に関わる端子、他は第1図に同様である。次に新たに加わった部分の動作を

説明する。

状パラメータ修正部7は、メモリ9に蓄えられた前フレームにおける口形状パラメータと、口形状パラメータ取得部5より与えられる現在の音素に対する口形状パラメータとに基づいて、例えば両者の中間値を求めて、現フレームの画像を合成するための口形状パラメータとする働きをする。スイッチ40は、遷移検出部から出力線210、211のいずれに信号が出力されるかによって、端子910或いは911に接続され、端子910に接続された時には、口形状パラメータ修正部7より得られる2つの音素に対する口形状パラメータの中間値を、また、端子911に接続された時には現在の音素に対する口形状パラメータを、画像生成部6に渡す。以上の例では、ある音素の口形状パラメータと次の音素の口形状パラメータとの中間値は1フレーム分しか生成されないが、例えばカウント82の値に応じて何段階かの中間値を生成することにより、より滑らかな口形状変化を実現することも可能である。

以上述べたように、本発明は文字列として表現

される文章を入力した場合にこれに対応した口形状変化を有する顔動画像を合成する方式に関するものである。しかしながら、音声情報を入力した場合においても入力音声情報に対してこれを音素列に分割し、各音素ごとに音声特徴及び持続時間を出力することが可能な音声認識手法が利用できるのであれば、本発明における音声合成部1をこのような動作をする音声認識部に置き換えることにより、入力音声情報に対応した口形状変化を有する顔動画像を合成することも可能である。

#### (発明の効果)

以上のように、本発明により、文字列として表現される文章を入力として音声出力との的確な対応付けがなされ、かつ、各音素の持続時間に合わせた口形状変化を有する、従って音声出力とのマッチングのとれた自然な口形状変化を有する動画像を合成することが可能である。

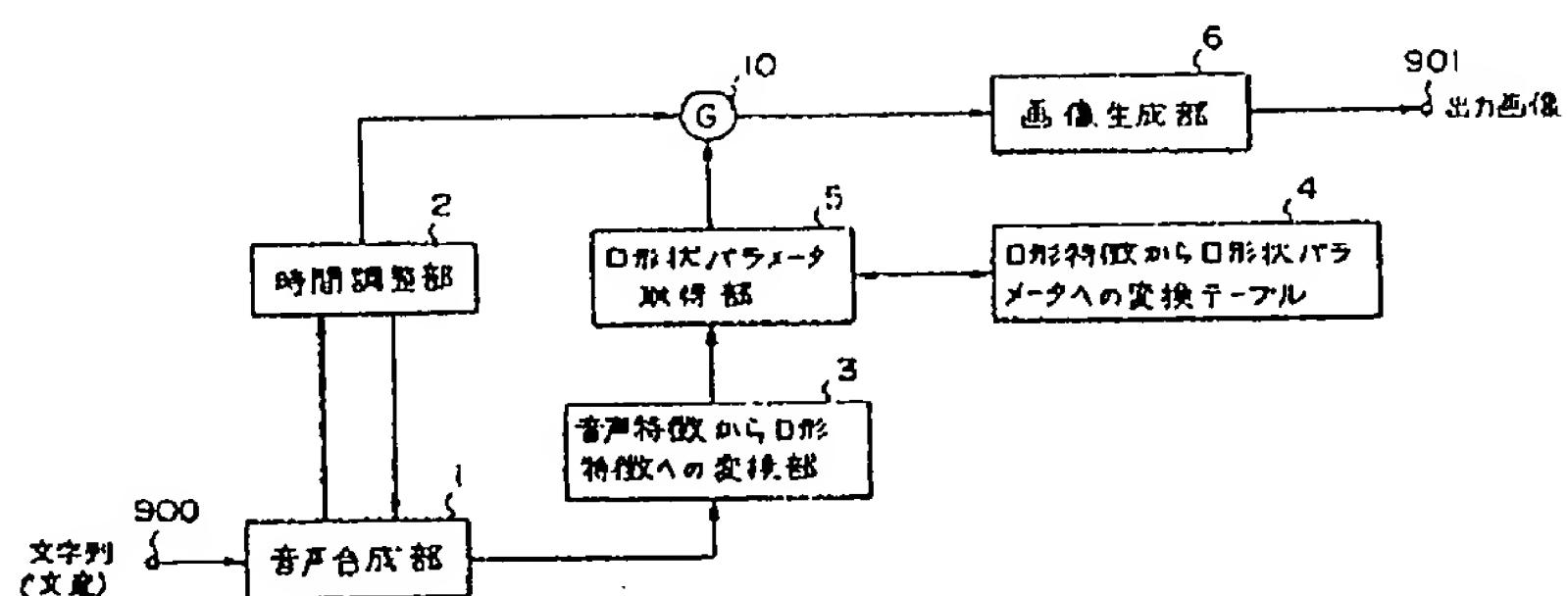
文章入力に対して、今まで音声を合成するのにとどまっていたのに対し、本発明では、音声とのマッチングのとれた自然な口形状変化を有する動

画像まで容易に出力できるようになる。従って、本発明は実写を必要とせずにリアルな動画像を生成する用途（例えば、放送番組や映画の製作）、音声及び画像による自動応答装置、マン・マシン・インターフェースの手段としての利用、文章から音声及び動画像へのメディア変換等に適用可能であり、その効果が極めて大である。

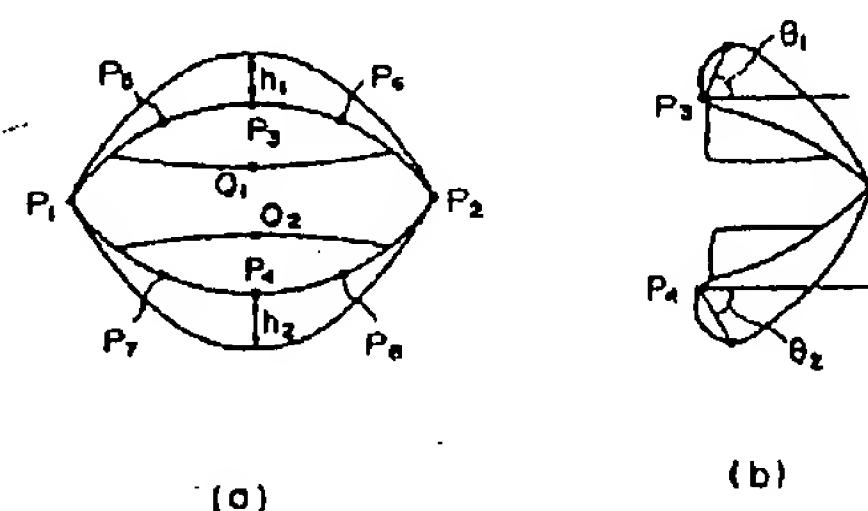
#### 4. 図面の簡単な説明

第1図は本発明の第1の実施例に対応するブロック図、第2図は口形状を表現するためのパラメータの例を示した図、第3図は本発明における時間調整部2の動作の一例に対応するブロック図、第4図は本発明の第2の実施例に対応するブロック図、第5図は本発明の第2の実施例における遷移検出部8の動作の一例に対応するブロック図、第6図は従来の画像合成方式の動作に対応するブロック図である。

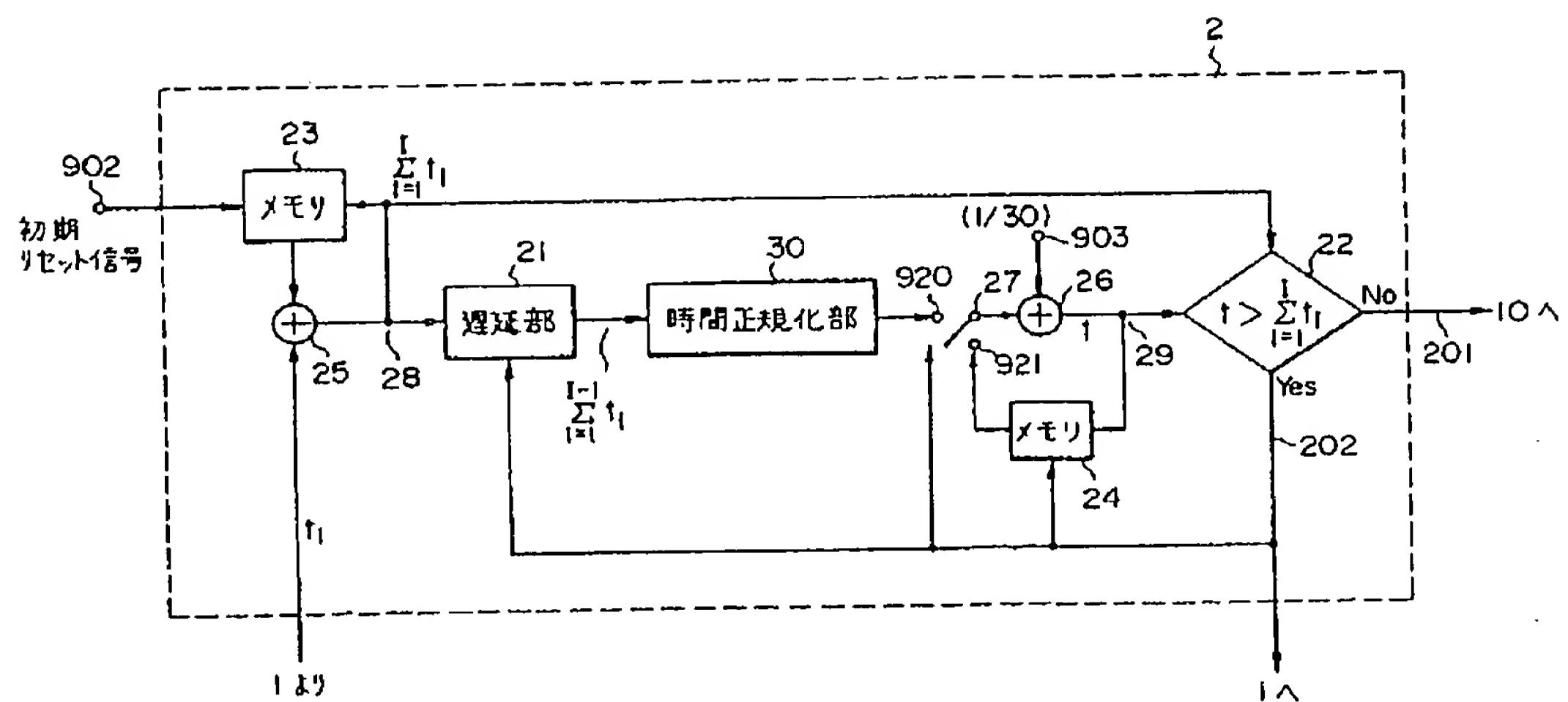
第1図



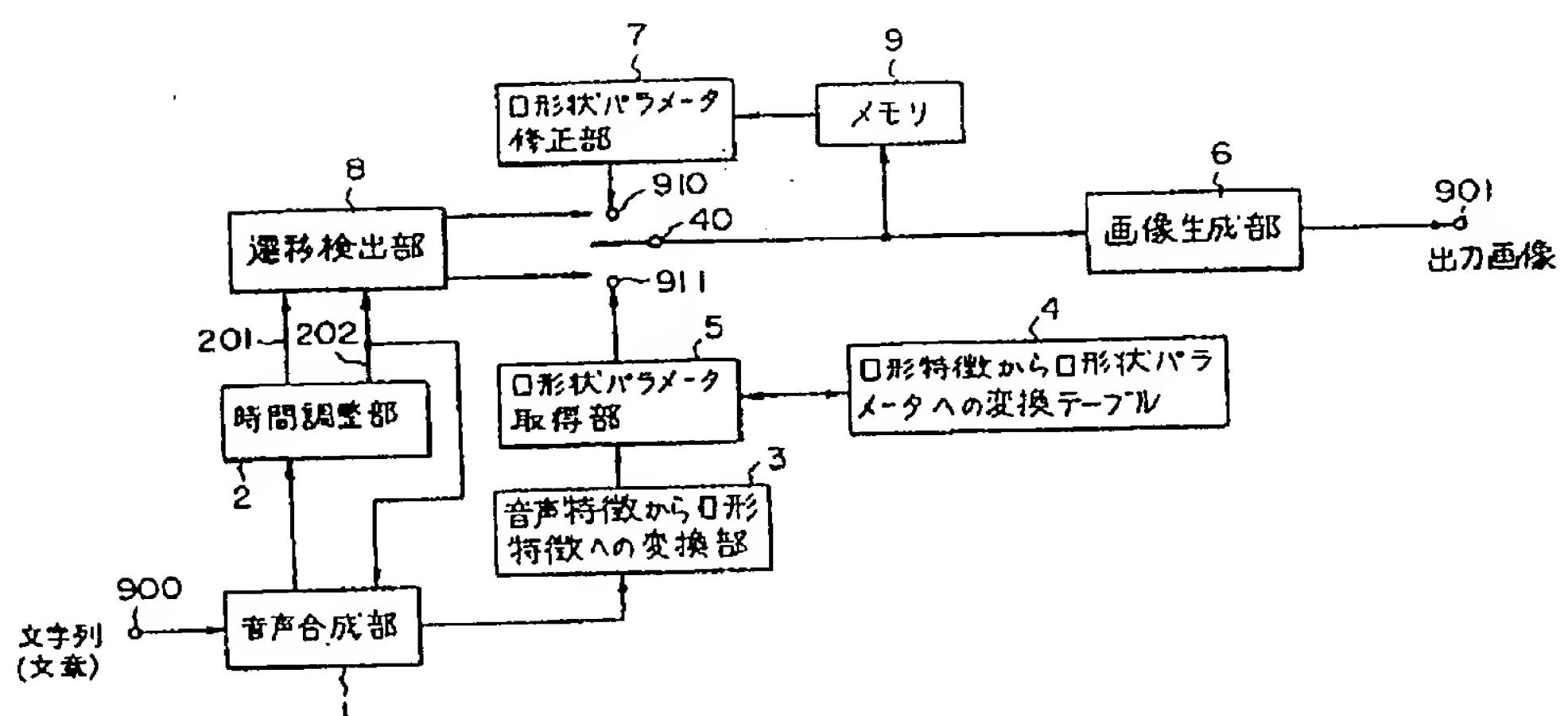
第2図



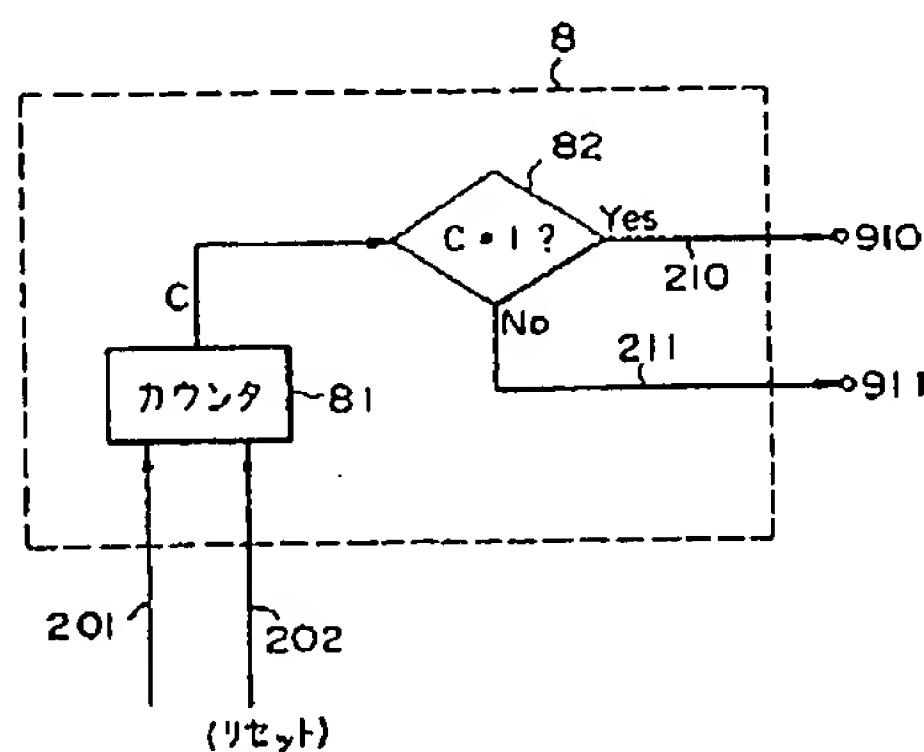
第3図



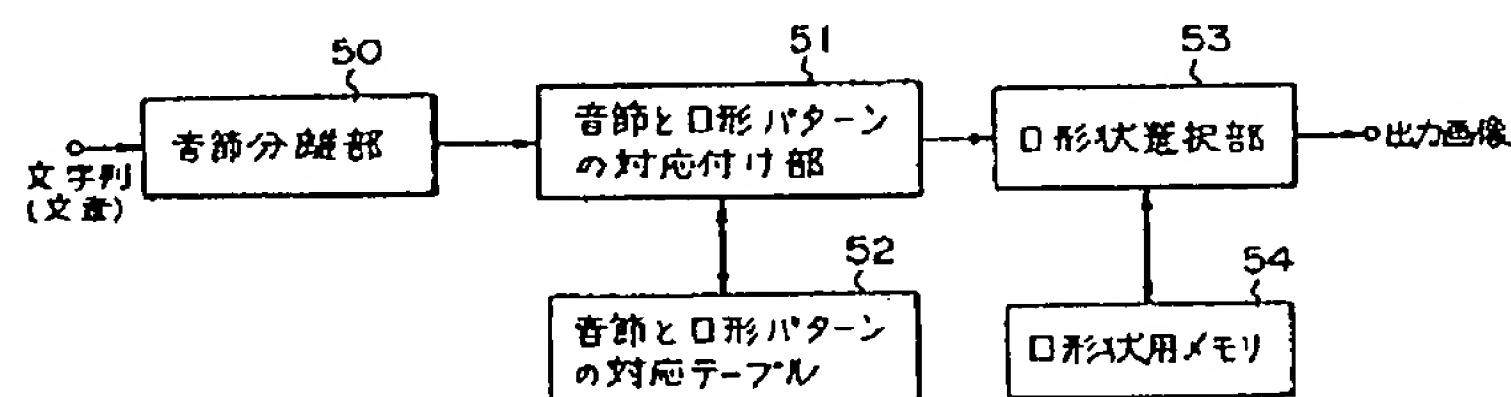
第4図



第5図



第6図



第1頁の続き

⑦発明者 橋口 宜男 東京都新宿区西新宿2丁目3番2号 国際電信電話株式会社内